Strategic Evolution of ESE Data Systems

Lifecycle Study Breakout

Matt Schwaller
NASA/GSFC
schwaller@gsfc.nasa.gov

SEEDS Second Public Workshop
June 17-19, 2002

Agenda

1:00 (30 minutes) Introductions, study team results (Schwaller)

…and review the agenda proposed here…

1:30 (90 minutes) Panel Comments + Discussion

- Jon Christopherson -- Spacecraft/Pre-launch
- Robert Wolfe -- Product Generation
- Steve Kempler -- Active Archive
- Ben Watkins -- LTA

3:00 (60 minutes) Break into 4 sub-groups to review & comment on the report

4:00 (60 minutes) Report comments/findings group by group

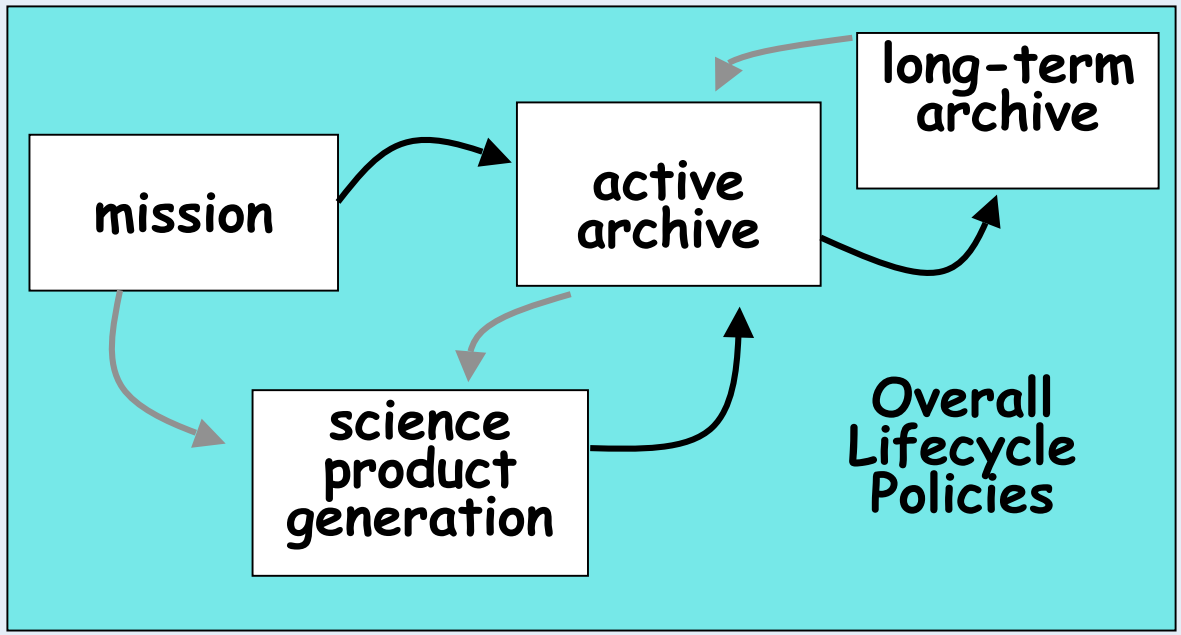
5:00 Final words & adjourn

Goals for this Afternoon

- **Ultimate goal: establish guidelines for ESE policies and procedures**
  - Incorporate lifecycle guidelines, policies and best practices into NASA Policy Directives, Announcements of Opportunity, NASA Research Announcements, etc.
- **Today: review the lifecycle study results so far (copies available)**
  - Provide comments and suggestions for improvements

Study Results-1

- Section 2: A simple model of four major lifecycle entities within a context of an overall set of guiding policies



- Recognizing that some of these functions may be grouped together in any given mission, and other inherent simplifications
- Guidelines are provided for each of these entities, for the (bolded) interfaces between them, and for "overall" polices

Study Results-2

- **Section 4: Overall guidelines**
  - There will be an **archive defined for each data product**
  - The **active archive and long-term archive will be identified at the beginning of the mission** (noting that for some products the active and long-term archive may be the same)
  - **All archive data collections will be complete** and will include all required ancillary data, project and data set documentation, and the science production software
  - **Data will be available** throughout its lifecycle **without loss or degradation in quality**
  - **Throughout a product's lifecycle a point-of-contact will be provided** that can be utilized for questions about the data or use of the data
  - Once a physical transfer of any data has occurred and formally accepted by the archiving site, **the data become the responsibility of the accepting party**
  - **Data may be transferred to other archives** such as the National Archives and Record Administration (NARA) when deemed appropriate.

Questions for Panelists & Participants

Lifecycle Model

- **Does it capture the essence of the ESE data lifecycle or does it need to be modified?**
- **What about the "missing" interfaces?**
  - The report argues that the "shaded"interfaces and interfaces to users are "enabled by successful implementation of the SEEDS product lifecycle that is studied" in the report. Is this argument believable?

Section 4: Overall Recommendations

- **The goal here is a set of high-level policy recommendations. Does anything need to be added, revised or removed?**

More Questions

Sections 5-7: Mission Responsibilities, Science Product Generation, Active Archive

- Includes both "internal" and "interface" recommendations
- Attempt was made to minimize "internal" recommendations (to let these entities define their own levels of service to the greatest extent possible)
- Overall, an attempt was made to keep the "cost" of these guidelines and recommendations to a minimum: is this the appropriate level of responsibility? Are the costs too high to be acceptable to a mission, SIP or archive? Do we need to add additional responsibilities to ensure successful lifecycle transitions?

Section 8: LTA

- Only "internal" recommendations
- This was the hardest set of recommendations to write, and probably needs the most work … comments welcome!

Next Steps

- Immediate: based on results of this workshop…
  - Refine and revise the SEEDS lifecycle study guidelines and recommendations
- Near term: integrate study results with other study results
  - Especially, parameterization of the SEEDS cost study
  - Contribute to and draw on results of reuse, metrics and standards studies
- Next year: continue study with NOAA participation as funds and interest allow
- Also next year: draft language for NPDs, AOs, etc.

Mission & Product Generation

- **Issues/Recommendations:**
- **"Data-buys" or any proprietary mission needs a time-period to be defined after which the data becomes the property of NASA**
- **Intellectual property rights issues need to be addressed**
  - PI institution needs to commit/agree to make all archive data products available (where possible) to the public (or a sunset clause needs to be defined)
  - Or, at least identify which tools and products were used, and identify how the data may be accessed
- **Relevance of data quality act? (Watermark, provenance issues, reproducibility of data, peer review, integrity of data and supporting information)**
- **Issue of keeping, discarding data: CERES doesn't throw anything away … MODIS does discard old versions**
- **Preservation of "community knowledge": need early involvement of mission and product generation communities in defining the content and LOS of the archive**
  - Involvement of these people in advisory groups
- **Need to record production flows, production histories, PGE-versions, etc.**
- **Cost model: need to parameterize LOS, levels of documentation -- what is the level of documentation necessary?**
  - Need to bring the lifecycle and cost teams together

NEXT:

- **Revise the guidelines document with …**
  - Inputs from the 4 groups
  - Revise scope, contents and possibilities in each area
  - Incorporate today's issues/recommendations